Survey on Traffic Redundancy and Elimination Approach for Reducing Cloud Bandwidth and Costs

V. Uday Kumar [1], V. Savithri [2]

[1] M.Tech, Assistant professor of Rajeev Gandhi Memorial College of Engineering & Technology, Nandyal, Andhra Pradesh, India.

[2] MCA, Rajeev Gandhi Memorial College of Engineering & Technology, Nandyal, Andhra Pradesh, India.

Abstract— Cloud computing is a fast growing field which is arguably a new computing paradigm. In cloud computing, computing resources are provided as services over the internet and users can access resources on based on their payments. Transmission cost plays an important role when trying to minimize cloud cost. However for server specific TRE approach it is difficult to handle the traffic efficiently and it doesn't suites for the cloud environment because of high processing costs. In this paper we give a survey on the new traffic redundancy technique known as novel-TRE also known as receiver based TRE. This novel-TRE has significant features like detecting the redundancy at the client, repeats appear in chains, matches incoming chunks with a previously received chunk chain or local file and sending to the server for predicting the future data and no need of server to continuously maintain client state.

Keywords: Cloud Computing, Lightweight chunking, pay-as-you-go, novel-TRE.

I. INTRODUCTION

Cloud computing is emerging style of delivery in which applications, data and resources are rapidly provisioned as standardized offerings to users with a flexible price. The cloud computing paradigm has achieved widespread adoption in recent years. Its success is due largely to customers' ability to use services on demand with a pay-as-you go [2] pricing model, which has proved convenient in many respects. Low costs and high flexibility make migrating to the cloud compelling. Cloud computing is the long dreamed vision of computing as a utility, where users can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. By data outsourcing, users can be relieved from the burden of local data storage and maintenance. Traffic redundancy and elimination approach is used for minimizing the cost.

Cloud applications that offer data management services are emerging. Such clouds support caching of data in order to provide quality query services. The users can query the cloud data, paying the price for the infrastructure they use. Cloud management necessitates an economy that manages the service of multiple users in an efficient, but also, resource economic way that allows for cloud profit. Naturally, the maximization of cloud profit given some guarantees for user satisfaction presumes an appropriate price-demand model that enables optimal pricing of query services. The model should be plausible in that it reflects the correlation of cache structures involved in the queries. Optimal pricing is achieved based on a dynamic pricing scheme that adapts to time changes.

Protocol-independent redundancy elimination [3] works on Manber to detect similar, but not necessarily identical, information transfers. In terms of improving Web performance, it has the potential to exceed the benefits of other approaches such as delta coding and duplicate suppression. This is because the similarity algorithms on which it is based include as a subset both exact matches (duplicate suppression) and differences between versions of the same document (delta coding).A distinguishing feature of our system is that it is protocol independent. It makes no assumptions about the syntax or semantics of HTTP. This has two distinct advantages. It is able to identify fine grained sharing, as may be common with dynamically generated or personalized pages, as well as inter-protocol sharing. It does not need to be updated to take advantage of new types of content, such as streaming media, as they emerge or delivery protocols are revised.

II. TRE ENVIRONMRNT IN CLOUD COMPUTING

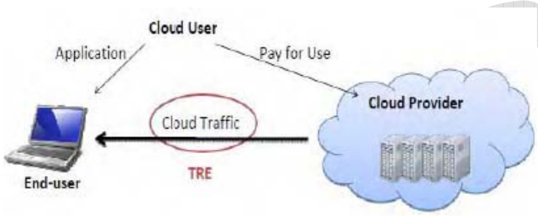

Fig: 1 Cloud TRE Environment

Fig 1 shows how the traffic encounters in the cloud computing between end user and the cloud server by transferring the same content repeatedly. In order to removing redundancy we are using TRE approach. Large amount of popular content is transferred repeatedly across network links in the internet. To transfer the information between the sender and receiver data is divided into chunks. Chunking mechanism helps to improve the efficiency by parallel upload/download of different chunks and each chunk generates a signature to transfer the data in a secure manner.

III. Wide Area Network TRE

Traffic happens between the sender and receiver by transmitting the same information repeatedly, that information may contain files, documents, video's etc. In recent years various TRE techniques have been developed for eliminating the redundant data. Wanax [4] present, Wide-area network (WAN) accelerators operate by compressing redundant network traffic from point-to-point communications, enabling higher effective bandwidth.

A.Multi Resolution Scheme:

Wanax uses a multi-resolution chunking (MRC) scheme that provides high compression rates and high disk performance for a variety of content, while using much less memory.

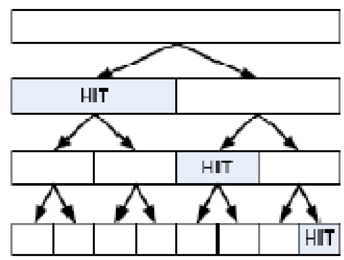

Fig 2: MRC Aligned Chunk Boundaries

Fig [2] shows MRC (multi resolution chunking) operation, here the received data is divided into equal sized chunks. Content fingerprinting (CF) forms the basis for Wanax, since it provides a position-independent and history-independent technique for breaking a stream of data into smaller pieces, or chunks, based only on their content.

To prevent chunks from being too large or too small, minimum and maximum chunk sizes can be specified as well. Since Rabin fingerprinting determines chunk boundaries by content, rather than offset, localized changes in the data stream only affect chunks that are near the changes. Once a stream has been chunked, the WAN accelerator can cache the chunks and pass references to previously cached chunks, regardless of their origin. Wanax approach is based on three-way-handshake technique uses a sender middle-box and receiver middle-box for transmitting data between the sender and the receiver.

B.Disadvantages:

(1) End-to- end encrypted traffic do not cope well middle-boxes.

(2) It creates latency for non cached data and middle-boxes will not improve the performance.

Fig [2] shows MRC (multi resolution chunking) operation, here the received data is divided into equal sized chunks. Content fingerprinting (CF) forms the basis for Wanax, since it provides a position-independent and history-independent technique for breaking a stream of data into smaller pieces, or chunks, based only on their content.

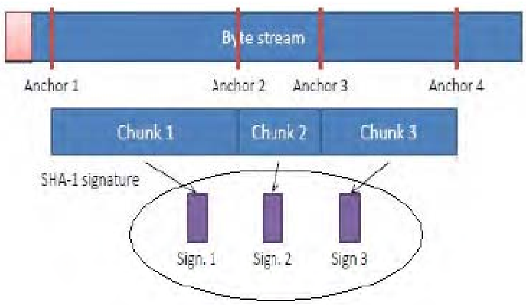

Fig 3: Lightweight Chunking

The novel-TRE receiver maintains a *chunk store*, which is a large size cache of chunks and their associated metadata. Chunk's metadata includes the chunk's signature and a (single) pointer to the successive chunk in the last received stream containing this chunk. When the new data are received and parsed to chunks, the receiver computes each chunk's signature using SHA-1. At this point, the chunk and its signature are added to the chunk store. In addition, the metadata of the previously received chunk in the same stream is updated to point to the current chunk..

B. Prediction Operation:

Fig-4 shows how the chunks are predicting in the receiver, upon the arrival of new data the receiver computes the respective signature for each chunk and looks for a match in its local chunk store.

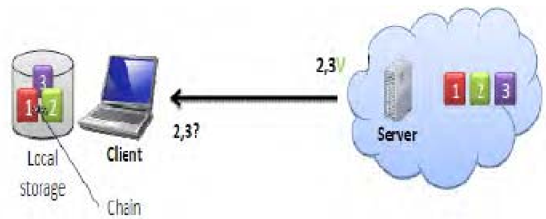

Fig 4: Prediction Operation

If the chunk's signature is found, the receiver determines whether it is a part of a formerly received chain, using the chunks' metadata. If affirmative, the receiver sends a prediction to the sender for several next expected chain chunks. Upon a successful prediction, the sender responds with a PRED-ACK confirmation message. Once the PRED-ACK message is received and processed, the receiver copies the corresponding data from the chunk store to its TCP input buffers, placing it according to the corresponding sequence numbers. At this point, the receiver sends a normal TCP ACK with the next expected TCP sequence number. In case the prediction is false, or one or more predicted chunks are already sent, the sender continues with normal operation, e.g., sending the raw data, without sending a PRED-ACK message. On the other hand, the use of smaller chunks increases the storage index size, memory usage, and magnetic disk seeks. It also increases the transmission overhead of the virtual data exchanged between the client and the server.

VI. CONCLUSION

In this paper, we proposed a novel-TRE approach for eliminating redundancy in the cloud environment. Our proposed scheme has significant features like reduces the transmission cost by predicting chunks, redundancy detection by the client, does not require the server to continuously maintain clients' status. Our receiver based end-to-end TRE suites for cloud environment.

V. REFERENCES

[1] E. Zohar, I. Cidon, and O. Mokryn, "PACK: Prediction-Based Cloud Bandwidth and Cost Reduction System. IEEE/ACM Transactions on Networking, 01 February 2013.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph,R.Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, 2010.

[3] N. T. Spring and D. Wetherall, "A protocol-independent technique for eliminating redundant network traffic," in Proc. SIGCOMM, 2000, vol. 30, pp. 87–95.

[4] Ihm, K. Park, and V. Pai, "Wide-area network acceleration for the developing world," in Proc. USENIX ATC, 2010, pp. 18–18.

[5] B. Aggarwal, A. Akella, A. Anand, A. Balachandran, P. Chitnis, C. Muthukrishnan, R. Ramjee, and G. Varghese, "EndRE: An end-system redundancy elimination service for enterprises," in Proc. NSDI, 2010,pp. 28–28.

[6] Ben-Shaul, "Method and apparatus for reducing network traffic over low bandwidth links," US Patent 7636767, Nov. 2009.

[7] Anand, A. Gupta, A. Akella, S. Seshan, and S. Shenker, "Packet caches on routers: The implications of universal redundant traffic elimination," in Proc. SIGCOMM, 2008, pp. 219–230.

[8] Z. Broder, "Some applications of Rabin's fingerprinting method, "in Sequences II: Methods in communications, Security, and Computer Science. New York, NY, USA: Springer-Verlag, 1993, pp. 143–152.

[9] U. Manber, "Finding similar files in a large file system," in Proc.USENIX Winter Tech. Conf., 1994, pp. 1–10.

BIOGRAPHY

Author 1: Details,

Name: V. Uday Kumar, B.Tech in Computer Science & Engineering M.Tech in Computer Science & Engineering, Working as Assistant Professor in RGM Engineering College, Nandyal

Author 2: Details,

Name: V. Savithri received MCA degree from Jawaharlal Nehru Technological University, Anantapur, Savithri766@gmail.com.